Journal of Dentistry 122 (2022) 104115

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

Review article

Deep learning for caries detection: A systematic review

Hossein Mohammad-Rahimi [a,b], Saeed Reza Motamedian [a,c], Mohammad Hossein Rohban [b], Joachim Krois [a,d], Sergio E. Uribe [a,e,f,g], Erfan Mahmoudinia [c], Rata Rokhshad [a], Mohadeseh Nadimi [h], Falk Schwendicke [a,d,*]

[a] Topic Group Dental Diagnostics and Digital Dentistry, ITU/WHO Focus Group AI on Health, Berlin, Germany
[b] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
[c] Dentofacial Deformities Research Center, Research Institute of Dental Sciences & Department of Orthodontics, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[d] Department of Oral Diagnostics, Digital Health and Health Services Research, Charité – Universitätsmedizin Berlin, Berlin, Germany
[e] Department of Conservative Dentistry and Oral Health & Bioinformatics Research Unit, Riga Stradins University, Riga, Latvia
[f] School of Dentistry, Universidad Austral de Chile, Valdivia, Chile
[g] Baltic Biomaterials Centre of Excellence, Headquarters at Riga Technical University, Riga, Latvia
[h] Cardiovascular Diseases Research Center, Department of Cardiology, Heshmat Hospital, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Neural networks
Dental caries
Dentistry
Systematic review

ABSTRACT

Objectives: Detecting caries lesions is challenging for dentists, and deep learning models may help practitioners to increase accuracy and reliability. We aimed to systematically review deep learning studies on caries detection.
Data: We selected diagnostic accuracy studies that used deep learning models on dental imagery (including radiographs, photographs, optical coherence tomography images, near-infrared light transillumination images). The latest version of the quality assessment tool for diagnostic accuracy studies (QUADAS-2) tool was used for risk of bias assessment. Meta-analysis was not performed due to heterogeneity in the studies methods and their performance measurements.
Sources: Databases (Medline via PubMed, Google Scholar, Scopus, Embase) and a repository (ArXiv) were screened for publications published after 2010, without any limitation on language.
Study selection: From 252 potentially eligible references, 48 studies were assessed full-text and 42 included, using classification ($n = 26$), object detection ($n = 6$), or segmentation models ($n = 10$). A wide range of performance metrics was used; image, object or pixel accuracy ranged between 68%-99%. The minority of studies ($n = 11$) showed a low risk of biases in all domains, and 13 studies (31.0%) low risk for concerns regarding applicability. The accuracy of caries classification models varied, i.e. 71% to 96% on intra-oral photographs, 82% to 99.2% on peri-apical radiographs, 87.6% to 95.4% on bitewing radiographs, 68.0% to 78.0% on near-infrared transillumination images, 88.7% to 95.2% on optical coherence tomography images, and 86.1% to 96.1% on panoramic radiographs. Pooled diagnostic odds ratios varied from 2.27 to 32,767. For detection and segmentation models, heterogeneity in reporting did not allow useful pooling.
Conclusion: An increasing number of studies investigated caries detection using deep learning, with a diverse types of architectures being employed. Reported accuracy seems promising, while study and reporting quality are currently low.
Clinical significance: Deep learning models can be considered as an assistant for decisions regarding the presence or absence of carious lesions.

1. Introduction

Artificial intelligence (AI) is defined as “the science and engineering of making intelligent machines” which can solve problems instead of humans [1]. Machine learning is one of the main subcategories of AI, enabling machines to learn and improve from experience without being explicitly programmed for a single task [2]. Typically, computers use example data that is extracted discriminant features, which are mostly

* Corresponding author.
E-mail address: falk.schwendicke@charite.de (F. Schwendicke).

https://doi.org/10.1016/j.jdent.2022.104115
Received 2 January 2022; Received in revised form 24 March 2022; Accepted 28 March 2022
Available online 30 March 2022

handcrafted, from images in order to train machine learning systems [3].

As a subset of machine learning models, deep learning models are built based on neural networks (NNs), which are biologically inspired programming architectures that allow computers to learn by observing patterns in data [4]. These networks are composed of many layers, which transform input data (such as raw images) into outputs (such as diagnoses) while learning higher-level features automatically [3]. Convolutional NN (CNN), which is a modification of NN, has proven most successful for the analysis of images to date. They are designed to integrate spatial information and configuration of both 2D and 3D images [5].

CNNs are widely used in medical applications, including drug development, diagnostics, treatment, and associated processes [6,7]. Deep learning has proven especially powerful for the analysis of complex data, like imagery [8], and has been used for image classification (i.e., labeling the image, e.g., signs of disease are present), detection (e.g. signs of disease are present in this area, usually indicated by a bounding box), and segmentation (i.e., signs of disease are present on these specific pixels). The accuracy of deep learning for medical image analysis has been found to match or, in some cases, surpass that of experts [9]. In dentistry, deep learning has been employed for image analysis in orthodontics [2], specifically landmark analysis on cephalometric radiographs [10], endodontics (detection of apical lesions), periodontology (periodontal bone loss) and cariology [9].

Dental caries is the most prevalent condition in the human population [11]. Caries lesions can be detected by visual and tactile means, while often this visual-tactile detection is supplemented by imaging strategies like radiography (the most common type of adjunct detection method), optical coherence tomography, quantitative light-induced fluorescence, intraoral scanner, or near-infrared light trans-illumination [12–14]. Evaluating any imagery for caries detection is a challenge for practitioners [15,16]; dentists miss a substantial proportion of early caries lesions in radiographs [17] but also other image types [18–20], and show considerable variability in their diagnostic findings and treatment decisions.

The usage of AI, specifically deep learning, may support practitioners in caries lesion detection and diagnosis on imagery [15,21–23]. Notably, the development in the field of deep learning of caries detection is highly dynamic; moreover, studies show substantial variability in methods and outcomes. We, therefore, aimed to systematically appraise studies using deep learning (specifically, NNs) for caries detection on dental imagery and compare their outcomes regarding reported performance measurements including accuracy, specificity and sensitivity. We did not further specify the task (image classification, detection, segmentation) or the image modality. We further aimed to synthesize our findings and to evaluate the robustness of overall body of evidence.

2. Materials and methods

This is a systematic review of diagnostic accuracy studies. Reporting of this study follows the PRISMA-DTA guideline [24]. The study protocol was registered at PROSPERO (CRD42019125491).

2.1. Eligibility criteria

This systematic review answers the following PICO question: What are the applications and accuracy (outcome) of deep learning (intervention) for caries detection on dental imagery (population) in comparison with the reference test (additionally and, if available, also against the standard of care, i.e. human experts without deep learning)?

Studies reporting the following criteria were included:

P: Studies used deep learning models on dental imagery resulting from routine care, clinical studies or studies on extracted teeth;

I, C: Caries classification, detection, or segmentation models with a deep learning (NN) architecture, compared with a reference test

O: Any kind of accuracy estimate on image, tooth, surface or pixel level.

Studies with the following criteria were excluded: Studies without sufficient details on the data used for training and testing (e.g., dataset size, data modality, *etc.*); studies without a clear explanation of the deep learning model; studies that did not separate the accuracy of caries detection with that of detecting other pathologies; reviews.

2.2. Information sources and search

An electronic search was conducted in the following electronic databases up to 12th April 2021: Medline (via PubMed), Google Scholar, Scopus, Embase, and ArXiv. Search results were limited to publications after 2010, accounting for the fact that deep learning for image analysis has been popularized in 2012 by Krizhevsky et al. [25]. There was no limitation in the language. Each database was searched with adapted keywords. The search query for each database is described in Table 1. Journal articles and conference proceedings were screened to identify further studies. Moreover, manual cross-referencing of the bibliographies of included papers was conducted.

2.3. Study selection

For managing the citations, Endnote X9 (Clarivate, Philadelphia, USA) was used. Two independent reviewers performed title and abstract screening after removing duplicate papers (M.N and R.R). Any disagreement was resolved through consensus with a third reviewer (H.M.R). Then, two independent investigators evaluated full texts of eligible studies based on inclusion and exclusion criteria (M.N and R.R). Any disagreements or discrepancies between the two reviewers were resolved through consensus by a third investigator (H.M.R).

Table 1
The results of the electronic search in the various databases.

Database	Keywords	Results	Date
PubMed	("Artificial Intelligence" OR artificial intelligence [MeSH Terms] OR "deep learning" OR "deep learning"[Mesh Terms] OR "machine learning" OR "machine learning"[MeSH Terms] OR "neural network" OR "computer vision") AND caries	59	12th April 2021
Google Scholar	allintitle:("artificial intelligence" OR "deep learning" OR "machine learning" OR "computer vision" OR "neural network") AND caries	28	12th April 2021
Scopus	TITLE-ABS-KEY (("Artificial Intelligence" OR "deep learning" OR "machine learning" OR "neural network" OR "computer vision") AND ("caries")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT- TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011) OR LIMIT-TO (PUBYEAR , 2010))	119	12th April 2021
Embase	('artificial intelligence':ti,ab,kw OR 'deep learning':ti,ab,kw OR 'machine learning':ti,ab,kw OR 'neural network':ti,ab,kw OR 'computer vision':ti,ab,kw) AND 'caries':ti,ab,kw AND [2010–2021]/py	42	12th April 2021
ArXiv	AND all="Artificial Intelligence" OR "deep learning" OR "machine learning" OR "neural network" OR "computer vision"; AND all=caries	4	12th April 2021

2.4. Data collection and extraction

Two reviewers (M.N. and E.M.) independently collected data from the included studies. A third reviewer (H.M.R.) revised the data collection for discrepancies and disagreements. The following data items were extracted: bibliographic details (name of authors and the year of publication), data modality, dataset size (train/valid/test, if given), inclusion and exclusion criteria on population and image level (if available), caries prevalence, labeling procedure (i.e. how the reference test was established), task (i.e. classification, detection, segmentation), pre-processing, augmentation, deep learning approach (NN architecture), loss function employed, hardware used, accuracy measures employed and resulting findings. If more than one NN architecture was used, we only reported on the most accurate one.

2.5. Risk of bias and applicability

Two reviewers (M.N. and E.M.) independently used the QUADAS-2 tool [26] for risk of bias assessment. The QUADAS-2 checklist contains four risk of bias domains, including data selection, index test, reference standard, and flow and timing. It also has three domains evaluating the applicability of a study on patient selection, index test, and reference standard. Any disagreements or discrepancies between the two reviewers were resolved by a third investigator (H.M.R.).

In "patient selection", limited information on the dataset being presented as well as unclear data-split strategies and resulting data leakage were considered to indicate high risk of bias. For "index test", indicators were poor reporting on test reproducibility, insufficient information about the model construction and lack of robustness analyses of the model. For "reference standard", indicators, were lack of information on reference standard definition and using only one examiner for establishing the reference test. Finally, in "flow and timing", indicators were employing different reference standards across the same study and inappropriate intervals between the index test and reference standard. In the case of concerns regarding the applicability of the studies, we looked for the dataset used, the deep learning model and its performance, and the annotation procedure match review question in each domain, respectively. Leading questions used in the present study are reported in Table 2.

2.6. Synthesis

Due to a wide range of specific study designs and accuracy measures employed, we limited the quantitative synthesis to classification studies (as here, accuracy is always on image level; in detection and segmentation studies, accuracy can be determined on multiple levels, with accuracy measures only being limitedly comparable).

As very few studies reported the number of true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs) samples, but reported sensitivity and specificity, we used the diagnostic odds ratios (DORs) as pooled outcome for classification studies, calculated as follows:

$$DOR = \frac{Sensitivity \times Specificity}{(1 - Sensitivity)\ X\ (1 - Specificity)}$$

As a result, we limited the quantitative analysis to studies that reported both the specificity and the sensitivity. We further employed stratification according to detection threshold, i.e., enamel, enamel and dentin combined, only dentin, or unclear threshold.

3. Results

3.1. Study selection and study characteristics

Of 252 identified studies, a total of 48 studies were assessed full-text. Eleven of these studies were excluded after full-text assessment, the reasons for the exclusion are shown in Table S1. In total, following the manual search, 42 studies were included; the number of studies per year increased over the observation period, and the type of imagery became more diverse with time (Fig. 1).

Table 2
Modified leading questions of QUADAS-2 for critical appraisal.

Domain	Leading Question
Patient selection	1. Were data imbalances (if there were any) addressed in the study? 2. Did the study avoid inappropriate data exclusion? (e.g., difficult to diagnostic or considered as outliers without a specific outlier detection method) 3. Was a consecutive or random sample of patients (or data) enrolled? 4. Was the test dataset separate from the training and validation datasets? **Applicability:** Are there concerns that the included data and setting did not match the review question?
Index test	1. Were the deep learning method results interpreted without knowledge of the results of the reference standard? 2. If a threshold was used, was it pre-specified? 3. Was the method described in sufficient detail to reproduce the presented results? 4. Did the study perform any robustness or sensitivity analysis of their model? **Applicability:** Are there concerns that the method, its conduct, or interpretation differed from the review question?
Reference standard	1. Were the reference standard results interpreted without knowledge of the results of the index test? 2. Did the study use a gold standard? 3. If not, was the annotation procedure described in the study and found to minimize bias? 4. Did the study sufficiently report their limitations, biases, or issues around generalizability? **Applicability:** Are there concerns that the target condition as defined by the reference standard did not match the question?
Flow and timing	1. Were all data included in the analysis? 2. Was there an appropriate interval between the index test and reference standard? 3. Did all data have a reference standard? 4. Did all data have the exact reference standard?

Individual studies are summarized in Tables 3, 4, and 5, with each table showing studies using different deep learning tasks (classification, detection, segmentation). Overall, seven different types of imagery were employed; intra-oral photographs ($n = 12$), peri-apical radiographs ($n = 12$), bitewing radiographs ($n = 8$), optical coherence tomography images ($n = 4$), panoramic radiographs ($n = 3$), near-infrared light trans-illumination images ($n = 2$), and cone-beam computed tomography ($n = 1$) (Fig. 2). Of the employed image datasets, three were publicly available; the international symposium on biomedical imaging (ISBI) 2015 challenge dataset of bitewing radiographs (used by $n = 4$ studies) [27], a dataset of peri-apical radiographs ($n = 6$ studies) [28], and a dataset of intra-oral photographs used in a Kaggle competition ($n = 1$) [29]. One study reported that they employed peri-apical radiographs from a Kaggle competition, too, but the manuscript did not give further details towards the dataset.

Most studies used expert opinions to set the reference test ($n = 37$). Specifically, one human expert ($n = 8$), two ($n = 7$), three or more ($n = 6$) experts were involved in defining the reference test; 16 studies did not mention the number of experts involved. Two studies used other image modalities for establishing the reference test, namely fluorescent imaging [30] and micro-CT [31]. Four studies did not mention how the reference test was established.

Regarding the deep learning task, the most often chosen task was classification ($n = 26$), followed by segmentation ($n = 10$) and object detection ($n = 6$). Various deep learning models were used (Fig. 2). In classification studies, most of the studies used customized CNN structures ($n = 8$), transfer learning models ($n = 8$), or multi-layer perceptron ($n = 7$). Regarding segmentation, auto-encoders were used (e.g., U-net) ($n = 8$). Regarding detection, one-stage object detectors (e.g., YoLo) or

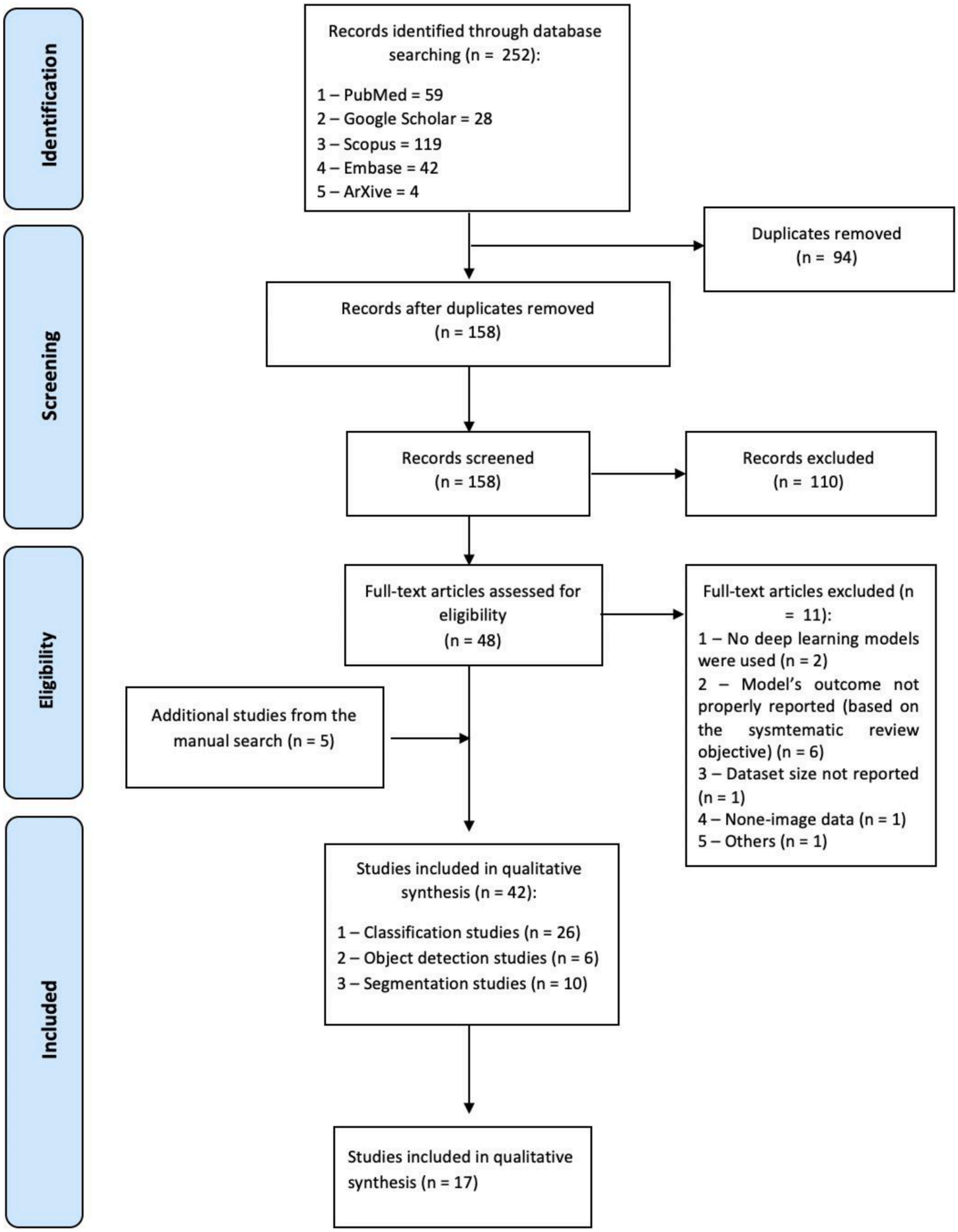

Fig. 1. . Flowchart of the search.

two-stage object detectors (e.g., Faster R-CNN) were used in the majority of studies ($n = 4$).

Classification studies by large reported on accuracy, sensitivity and specificity; other outcome measures were the F1-score, precision (positive predictive value) or the area-under-the receiver-operating-characteristics curve (AUC). In object detection and segmentation studies, outcome measures were more heterogeneous, and in addition to the mentioned ones, the mean Intersection over Union (mIoU) and the Dice coefficient index were employed (Tables 3–5).

3.2. Risk of bias and applicability

Detailed information about risk of bias and concerns regarding applicability are presented in Table S2 and Figure S2. Among the included studies, 11 (26.2%) were found to have low risk of biases in all four domains. Moreover, 13 studies (31.0%) were evaluated as low risk for concerns regarding applicability. The most problematic domain was "Reference Standard", where only 20 studies (47.6%) and 22 studies (52.4%) were classified as low risk of bias and low risk of applicability concern, respectively.

Table 3
Summary of findings in the selected classification studies.

Author, Year (Ref)	Data Modality	Dataset Size (Train/ Valid/Test)*	Inclusion & Exclusion Criteria (if any)	Caries prevalence in the (test) dataset	Labeling Procedure	Pre-Processing	Augmentations	Model Structure	Loss Function	Hardware	Performance measure (for caries detection, mean or single value)	Outcome
Leo 2021 [44]	Bitewing radiographs (from ISBI2015 [27])	80 bitewings, 480 tooth segments (300/180)	*NA*	*NA*	By 2 experienced dentists, 7 dental structures were annotated: caries, enamel, dentin, pulp, crown, restoration, root canal treatment. Based on the annotations, data were classified into 2 classes: healthy and carious teeth	Manual ROI extraction by cropping, Normalization, Selective median filter, Segmentation, Feature extraction	*NA*	Hybrid Neural Network (stacked sparse autoencoder + MLP)	*NA*	*NA*	Accuracy	0.96
Leo 2020 [45]	Bitewing radiographs (from ISBI2015 [27])	120 bitewings, 418 tooth segments (280/138)	*NA*	All: 56.70% Test: 47.83%	By 2 experienced dentists, 7 dental structures were annotated: caries, enamel, dentin, pulp, crown, restoration, root canal treatment. Based on the study's objective, extracted teeth were classified as healthy and carious teeth	ROI extraction (tooth crown), Grayscale, resize, Selective median filter	*NA*	Google Net inception v3	*NA*	*NA*	Accuracy	0.876
Tripathi 2019 [46]	Bitewing radiographs	800	*NA*	*NA*	*NA*	Extracting features using Local Binary Patterns	*NA*	MLP	*NA*	*NA*	Accuracy MSE	0.954 0.001
Sonavane 2021 [47]	Intra-oral photographs [29]	74 (48/12/ 14)	*NA*	All: 74.32% Test: 71.43%	Data were manually annotated into 2 classes: healthy and carious teeth	*NA*	Horizontal flip, Zoom	CNN	Binary cross-entropy loss	*NA*	Accuracy	0.714
Singh 2021 [48]	Intra-oral photographs	1500 (1200/ 300)	*NA*	*NA*	Data were manually annotated into 5 classes based on GV Black's classification + healthy teeth	Median filter for noise removal, Segmentation	*NA*	CNN-LSTM	Categorical cross-entropy	Intel Core (TM) i3 Lenovo platform with 3 GHz main processor speed and 8 GB memory	Accuracy Sensitivity Specificity Precision F1-score G-mean AUC	0.96 0.96 0.93 0.95 0.95 0.94 0.96
Wang 2020 [30]	Intra-oral photographs	7200 (6000/ 1200)	*NA*	*NA*	Classified by more than two dentists;	*NA*	*NA*	T-Net CNN	*NA*	*NA*	Accuracy Specificity Sensitivity	0.944 0.972 0.962

(*continued on next page*)

Table 3 (*continued*)

					Verification of the results was performed using the dual imaging system in both white and fluorescent modes. Data were annotated into 4 classes: Healthy, White Spot Lesion, Smashed, Plaque							
Singh 2020 [49]	Intra-oral photographs	400	*NA*	100%	Data were manually annotated into 5 classes based on GV Black's classification + healthy teeth	Feature extraction using GIST descriptor, Dimension reduction, Feature selection	*NA*	Probabilistic neural network	*NA*	*NA*	Accuracy Sensitivity Specificity	0.89 0.92 0.90
Guijarro-Rodríguez 2020 [50]	Intra-oral photographs	2030	*NA*	*NA*	*NA*	Gaussian filter, Median filter, Segmentation with Sobel filter, Feature extraction, PCA	*NA*	MLP	*NA*	*NA*	Accuracy	0.80
Prajapati 2017 [51]	Intra-oral photographs	251 (180/45/26)	*NA*	All: 31.97% Test: 30.76%	By dentists and oral and maxillofacial radiologists, data were annotated into 3 classes: Caries, periapical infection, periodontitis	*NA*	*NA*	VGG16	*NA*	*NA*	Accuracy	0.8846–0.875
Holtkamp 2021 [40]	Near-infrared light transillumination images	226 ex vivo 1319 in vivo* 5-fold cross-validation* trained and tested on both datasets.	Inclusion: Posterior teeth	NA	Three expert dentists independently annotated images pixel-wise and the fourth expert dentist reviewed all annotated images. Based on the union of all annotated areas on the image, the reference test was constructed.	Feature extraction	Horizontal and vertical flip, Zoom and Shift	ResNet	Binary cross-entropy loss	NVIDIA Quadro RTX 6000 GPU	Accuracy F1-score sensitivity specificity PPV NPV AUC	0.78 0.73 0.76 0.79 0.70 0.84 0.78
Schwendicke 2020 [18]	Near-infrared light transillumination images	226* 10-fold cross-validation	Inclusion: Posterior teeth	All: 40.71%	Two experienced dentists marked occlusal and/or proximal caries lesions pixel-wise, independently (following calibration	*NA*	Random resize, Random rotations, Horizontal and vertical flip	Resnet18, Resnext50	*NA*	*NA*	AUC (Resnet18-Resnext50) Accuracy (Resnet18-Resnext50)	0.73–0.74 0.69–0.68 0.46–0.59

(*continued on next page*)

Table 3 (*continued*)

					procedure). Data were annotated into 2 classes: teeth with or without carious lesions.						Sensitivity (Resnet18-Resnext50) Specificity (Resnet18-Resnext50) PPV (Resnet18-Resnext50) NPV (Resnet18-Resnext50)	 0.85–0.76 0.71–0.63 0.69–0.73
Salehi 2021 [19]	Optical coherence tomography images	81 (60/21)	Inclusion: human extracted premolar and molar teeth	*NA*	Data were manually annotated into 3 classes: Non-carious teeth, teeth with enamel caries and teeth with dentin caries	*NA*	Crop, Flip and Rotation	CNN	*NA*	*NA*	Accuracy Sensitivity Specificity PPV NPV	0.95 0.92 0.95 0.92 0.95
Salehi 2020 [20]	Optical coherence tomography images	81 (60/21)	Inclusio: Human extracted premolar and molar teeth	*NA*	Data were manually annotated into 3 classes: Non-carious teeth, teeth with enamel caries and teeth with dentin caries	Normalization	Crop, Flip, and Rotation	CNN	*NA*	*NA*	Accuracy Sensitivity Specificity PPV NPV	0.8873 0.7578 0.9621 0.9203 0.8731
Yu-Ping 2020 [31]	Optical coherence tomography images	748 cross-sectional images from 63 extracted teeth (599/149)	*NA*	All: 36.63%	Micro-CT was used for the reference standard; Data were annotated into 3 classes: Non-carious teeth, teeth with enamel caries and teeth with dentin caries	*NA*	*NA*	ResNet-152	*NA*	*NA*	Accuracy Sensitivity Specificity PPV NPV	0.9521 0.9885 0.8983 0.9348 0.9815
Salehi 2019 [52]	Optical coherence tomography images	51* 5-fold cross-validation	Inclusion: Human permanent teeth	All: 66.67%	Data were manually annotated into 3 classes: Non-carious teeth, teeth with enamel caries and teeth with dentin caries	*NA*	*NA*	CNN	*NA*	*NA*	Accuracy Sensitivity Specificity PPV	0.91 0.9846 0.998 0.999
Riyadi 2020 [53]	PA radiographs	237 radiographs, 660 cropped segments* 10-fold cross-validation	Inclusion: Teeth treated by pulp capping approach	*NA*	*NA*	Manual ROI extraction (tertiary dentine), Edge cropping	*NA*	CNN	*NA*	*NA*	Accuracy	0.91
Geetha 2020 [54]	PA radiographs	105* 10-fold cross-validation	*NA*	All: 46.68%	By a clinician, data were manually annotated into 2	Laplacian filter, Gaussian filter. Dilate and erode, Statistical	*NA*	MLP	*NA*	*NA*	Precision Recall F-measure MCC	0.963 0.962 0.962 0.924

(*continued on next page*)

Table 3 (*continued*)

					classes: healthy and carious teeth	feature extraction					AUC	0.983
Patil 2019 [55]	PA radiographs [28]	120 (84/36)	*NA*	*NA*	By a dental expert	Contrast enhancement, grey thresholding, Active contour	*NA*	MNP-ADA (MPCA + MLP)	*NA*	*NA*	Accuracy Sensitivity Specificity Precision F1-score MCC	0.95 1 0.5 0.95 0.97 0.69
Sornam 2019 [56]	PA radiographs [28]	120	*NA*	*NA*	By a dental examiner, other than healthy .vs carious teeth classification, data were labeled into 5 classes based on caries severity	Segmentation, Gray-level co-occurrence matrix feature extraction	*NA*	MLP	*NA*	*NA*	Accuracy MSE	0.99 0.003
Patil 2018 [57]	PA radiographs [28]	120 (80/40)* 3-fold cross-validation	*NA*	*NA*	By a dental examiner	Contrast enhancement, Gray threshold, Active contour	*NA*	MPCA + Customized CNN	L1	*NA*	Accuracy Sensitivity Specificity Precision	0.9 0.93 0.6 0.943
Patil 2018 [58]	PA radiographs [28]	120 (80/40)* 3-fold cross-validation	*NA*	*NA*	By a dental examiner	Contrast enhancement, Gray threshold, Active contour	*NA*	MPCA + MLP	L1 loss	*NA*	Accuracy Sensitivity Specificity	0.95 0.967 0.75
Lee 2018 [59]	PA radiographs	3000 (2400/600)	Exclusion: deciduous teeth	All & Test: 50%	By four calibrated board-certified dentists	Manual ROI extraction by cropping, Vertical flip	Rotation, Shear, Zoom, Width and Height shift, Horizontal flip	GoogleNet Inception V3	*NA*	*NA*	Accuracy Sensitivity Specificity MSE	0.82 0.81 0.83 0.003
Sornam 2018 [60]	PA radiographs [28]	120	*NA*	*NA*	By a dental examiner	Size Reduction, cropping, Smoothening, Sharpening and noise removal, GLCM feature extraction	*NA*	MLP	*NA*	*NA*	Accuracy MSE	0.991 0.008
Lakshmi 2021 [61]	Panoramic radiographs	1900 (1000/900)	*NA*	*NA*	*NA*	Histogram equalization, extracting the GLCM features	*NA*	AlexNet	*NA*	*NA*	Accuracy Precision Recall F1 score	0.960 0.785 0.842 0.982
Vinayahalingam 2021 [62]	Panoramic radiographs	500 segments of the third molar region (320/80/100)	Inclusion: Third molars, minimum age of 16, presence of at least one third molar; Exclusion: Blurred and incomplete panoramics	All: 50%	By medical records, then images were revalidated by two clinicians. If there was any disagreements, images were excluded.	Manual ROI extraction by cropping, Histogram equalization	Applied (augmentations approaches were not mentioned	MobileNet V2	*NA*	*NA*	Accuracy Precision F1-score Recall Specificity	0.87 0.87 0.87 0.87 0.86
Haghanifar 2020 [63]	Panoramic radiographs	1838 extracted tooth segments (80%/20%)	*NA*	All: 17.36%	Extracted teeth were classified as healthy and carious teeth by an oral and	Re-sampling of the smaller class, Vertical edge filter, Gaussian and bilateral	Horizontal flip, Vertical flip, Rotation, Brightness,	PaXNet (CNN + CapsuleNet)	*NA*	*NA*	Accuracy Precision Recall F0.5-score	0.860 0.894 0.506 0.78

(*continued on next page*)

Table 3 (*continued*)

maxillofacial radiologist. Mild and severe dental caries are also categorized. Annotation was done by an oral and maxillofacial radiologist.	filter, Sauvola binarization, ROI extraction, CLAHE, Feature extraction	Zoom, Width and Height shift

Abbreviations: **ADA,** Adaptive Dragonfly Optimizer; **AUC,** Area Under Curve; **CLACHE,** Contrast Limited Adaptive Histogram Equalization; **CNN,** Convolutional Neural Network; **GLCM,** Gray Level Co-Occurrence Matrix; **GPU,** Graphics Processing Unit; **LSTM,** Long Short-Term Memory; **MCC,** Matthews Correlation Coefficient; **MLP,** Multi-Layer Perceptron; **MPCA,** Modified Principal Component Analysis; **MSE,** Mean Squared Error; **NA,** Not Available; **NPV,** Negative Predictive Values; **PA,** Peri-Apical; **PPV,** Positive Predictive Values; **ROI,** Region of Interest.

* If only two numbers were written in this column usaully it means Train/Test splites.

3.3. Findings of the studies

The accuracy of deep learning for classifying images according to their caries status varied widely, i.e. 71% to 96% on intra-oral photographs, 82% to 99.2% on peri-apical radiographs, 87.6% to 95.4% on bitewing radiographs, 68.0% to 78.0% on near-infrared transillumination images, 88.7% to 95.2% on optical coherence tomography images, and 86.1% to 96.1% on panoramic radiographs.

As outlined, only classification studies were used for further synthesis. Of these, 17 studies could be pooled; the range of the reported sensitivity was 25% to 99.8% and specificity 17.2% to 100%, respectively. The majority of studies used both enamel and dentin lesions (i.e., a mixed threshold) or did not specify the threshold at all; those which defined the involvement as dentin as threshold found high sensitivity (above 90%) but mixed specificity (Fig. 3). For enamel as threshold, the opposite was true; studies found high specificity (above 95%) but mixed sensitivity. For combined or other thresholds, no clear pattern evolved. Similarly, the training dataset size was not clearly associated with sensitivity or specificity. DORs varied from 2.27 to 32,767; further analyses are displayed in Figure S2.

4. Discussion

The detection of early caries on imagery is challenging and marred by low sensitivity and inter-observer agreement [17]. Deep learning has been suggested to possibly assist in overcoming this problem, and in this systematic review, we systematically compiled and appraised studies involving deep learning for caries detection on images. We found a growing and increasingly promising body of evidence supporting deep learning for this task. Notably, though, studies were by large of limited quality and comparison across studies impeded by heterogeneity in conduct and reporting of the studies. A number of findings require more detailed discussion.

First, we found most studies to show relatively high accuracy, sensitivity, and specificity of deep learning for caries detection (usually exceeding 80%). Moreover, the few studies involving human experts as control group (and not only to establish the reference test) found deep learning similar or more accurate than humans. Generally, when considering the low accuracy of dentists especially for detecting early lesions [17], deep learning will likely yield useful accuracies for this task. Second, heterogeneity in study conduct and reporting was significant, impeding further comparisons across studies or quantitative synthesis. For example, studies employed a wide range of reference tests, with unclear impact on accuracy and bias. Particularly, a reference test defined by a single clinician only seems to come with serious bias, as any deep learning model will be as accurate (or worse) than this clinician in caries detection. Future studies are needed to (1) elaborate which strategies for establishing the reference test are most likely to yield the “truth” (e.g. majority vote, stepwise annotation, triangulation with second imagery), and (2) employ comprehensive and accurate data labeling to yield robust models [32]. Third, we had planned meta-analysis of accuracy outcomes if possible. For diagnostic accuracy studies, hierarchical models are recommended for such quantitative synthesis, but require the proportion TPs, TNs, FPs, and FNs as input data [33]. Since these values were reported by only a small minority of studies, this option was not available, though. Alternatively, separate pooling of sensitivity and specificity may be considered, but relies on the (unlikely) assumption that sensitivity and specificity are independent. Notably, this approach is also not recommended if studies used different diagnostic thresholds, as was the case here [33,34], and requires estimates of variance of both measures (which, again, were unavailable in most studies). Hence, we eventually decided to compute summary DORs as a single parameter of accuracy [35]. However, these are hard to interpret by clinicians [33], and we only pooled classification studies assuming that DORs of detection or segmentation models are not meaningful from a clinical perspective (but also to account for the fact that for these

Table 4
Summary of findings in the selected object detection studies.

Author, Year (Ref)	Data Modality	Dataset Size (Train/ Valid/ Test)*	Inclusion & Exclusion Criteria (if any)	Caries prevalence in the (test) dataset	Labeling Procedure	Pre-Processing	Augmentations	Model Structure	Loss Function	Hardware	Performance Measurements	Outcome
Srivastav 2017 [23]	Bitewing radiographs	3000 (2500/ 500)	*NA*	*NA*	By various dentists following clinical verification, annotations were as loose polygons around caries.	*NA*	*NA*	CNN	*NA*	*NA*	Recall Precision F1-Score	0.805 0.615 0.70
Zhang 2020 [64]	Intra-oral photographs	3932 (2507/ 300/ 1125)	None	All: 33.80% Test: 35.20%	By three board-certified dentists, each image annotated by a dentist independently with bounding boxes.	*NA*	Rotation, Hue, Saturation, and Exposure change	Single-shot detector	Smooth L1 (Localization) cross-entropy loss (Classification)	*NA*	AUC Image wise sensitivity Box-wise sensitivity	0.856 0.819 0.646
Javid 2020 [65]	Intra-oral photographs	90 (65/25)	*NA*	*NA*	By supervision of a dentist	Sharpening filter	*NA*	Mask RCNN (Based on ResNet50)	*NA*	*NA*	Precision Recall	0.95 0.96
Yu 2020 [66]	Intra-oral photographs	1368 (1052/ 316)	First permanent molar present	*NA*	By board-certified dentists	*NA*	Horizontal and vertical flip	ResNet-FPN, and two parallel task-specific subnetworks for region regression and region classification	Weighted Cross-entropy	Nvidia Titan XP GPU with 12 GB memory	Accuracy Sensitivity Specificity	0.952 0.898 0.961
Zhang 2020 [67]	Intra-oral photographs	1000 (800/ 200)	Inclusion: Anterior teeth	53.70%	By dental practitioners, sample, annotation classes: normal teeth, 6 classes of caries based on ICDAS classification system, others (crown, bracket, *etc.*)	*NA*	*NA*	Faster R-CNN (Based on ResNet50)	*NA*	Tesla K80 GPU	mAP AP50 AP75 AR	0.473 0.665 0.543 0.716
Choi 2016 [68]	PA radiographs	475 (380/ 95)* 5-fold cross-validation	*NA*	*NA*	Proximal caries was confirmed in periapical images by experts	Horizontal alignment, Cropping above the tooth top line and one-third of the image height from the bottom	Rotation, Left-right direction flip	CNN	*NA*	*NA*	F1-max FPs	0.74 0.88

Abbreviations: **AUC,** Area Under Curve; **CNN,** Convolutional Neural Network; **ICDAS,** International Caries Detection and Assessment System; **FP,** False Positive; **FPN,** Feature Pyramid Networks; **GPU,** Graphics Processing Unit; **L1 loss,** Least Absolute Deviations; **NA,** Not Available; **PA,** Peri-Apical; **R-CNN,** Region Based Convolutional Neural Networks; **ROI,** Region of Interest.

* If only two numbers were written in this column usaully it means Train/Test splites.

Table 5
- Summary of findings in the selected segmentation studies.

Author, Year (Ref)	Data Modality	Dataset Size (Train/Valid/Test)*	Inclusion & Exclusion Criteria (if any)	Caries prevalence in the (test) dataset	Labeling Procedure	Pre-Processing	Augmentations	Model Structure	Loss Function	Hardware	Performance Measurements	Outcome
Cantu 2020 [69]	Bitewing radiographs	3686 (3293/252/141)	Inclusion: Permanent teeth, crowns of min. one arch visible	*NA*	Three expert dentists independently annotated images pixel-wise and the fourth expert dentist reviewed all annotated images. Based on the union of all annotated areas on the image, the reference test was constructed.	*NA*	Flip, Center crop, Translation, Rotations, Gaussian-blur, Sharpening, Contrast, and Brightness	U-Net	Binary focal type	GeForce GTX 1080 Ti GPU using CUDA 10	Accuracy Sensitivity Specificity F1-score MCC	0.80 0.75 0.80 0.73 0.57
Kumar 2019 [70]	Bitewing radiographs	6000 (5000/1000)	*NA*	*NA*	By various dentists following clinical verification	*NA*	*NA*	U-Net	Cross-entropy loss	*NA*	Recall Precision F1-score	0.73 0.53 0.6142
Yun 2018 [71]	Bitewing radiographs [27])	120 (40/40/40)	*NA*	*NA*	By 2 experienced dentists, 7 dental structures were annotated: caries, enamel, dentin, pulp, crown, restoration, root canal treatment	*NA*	Rotation, Horizontal and vertical flip, Translation Transformation, Change of image gray value	Conditional generative adversarial network + U-Net	Arg-min(G) max(D) [cGAN(GD) + Lambda*L1 (G)]	*NA*	Precision True positives True negatives Dice similarity	0.546 – 0.418 0.784 – 0.768 0.956 – 0.973 0.697 – 0.584
Ronneberger 2015 [72]	Bitewing radiographs [27])	80 (40/40)	*NA*	*NA*	By 2 experienced dentists, 7 dental structures were annotated: caries, enamel, dentin, pulp, crown, restoration, root canal treatment	*NA*	Translations Rotations, Deformation gray value variations, (Increased to 20,000 training images)	U-net	Softmax loss	NVidia Titan GPU	Precision True positives True negatives Dice Similarity	0.453 0.576 0.983 0.564
Ezhov 2021 [73]	CBCT	4398	*NA*	*NA*	By dentists and oral and maxillofacial radiologists, then images were revalidated by the lead oral and maxillofacial radiologist, Labels were background, caries, metallic artifact and non-contrast filling.	ROI localization, tooth localization and numeration	*NA*	U-net	*NA*	*NA*	Sensitivity Specificity	0.7285 0.9953
Moutselos 2019 [74]	Intra-oral photographs	88 (79/9)	Inclusion: Posterior permanent molar Exclusion: Teeth	All: 100%	Occlusal surfaces were annotated pixel-wise into 7 classes: free of	Superpixel segmentation	Horizontal and vertical flip, Rotations, Shear	Mask R-CNN	*NA*	*NA*	micro F-measure	0.596, 0.625, 0.684

(*continued on next page*)

Table 5 (*continued*)

			with hypoplastic, hypomineralised areas and with sealants		caries and 6 classes of caries based on the ICDAS II classification system		and piecewise affine					
Casalegno 2019 [75]	Near-infrared light transillumination images	217 (185/ 32)	Exclusion: Images with dental restorations	*NA*	The overall agreement of 2 segmentation maps through manual labeling by experts; 5 dental structures were annotated: background, enamel, dentin, proximal caries, and occlusal caries	*NA*	Horizontal and vertical flip, Zoom, Rotation, Translation, Random contrast and brightness	U-net + VGG16	*NA*	*NA*	AUC mIOU	0.836 – 0.856 0.727
Khan 2021 [76]	PA radiographs	206 (132/44/ 30)* 4-fold cross-validation	Inclusion: Images used parallel technique, Exclusion: Images if at least 2 of 3 examiners did not agree	76.70% (In train + valid set)	By 3 experienced dentists. One examiner annotated data. The remaining 2 examiners evaluated the labels. Data were annotated into 3 classes: Caries, alveolar bone recession, and interradicular radiolucencies	*NA*	Magnification, Horizontal and vertical flip, Translation, Rotation	U-Net + Densenet121	*NA*	*NA*	mIoU Dice coefficient (all) mIoU Dice coefficient (caries)	0.383 0.434 0.194 0.239
Jung 2020 [77]	PA radiographs (from Kaggle)	86 (70%/ 30%)	*NA*	*NA*	By oral and maxillofacial radiologist, Data were manually annotated into 6 classes: apical problem abrasion, caries, impaction, periodontal problem, sound tissue.	*NA*	*NA*	DeepLab-v3+ (An autoencoder based on ResNet18)	*NA*	GeForce RTX 2070 GPUs, 32.0 GB RAM, AMD Ryzen 7 3800X 8-Core Processor (3893 Mhz, 8 core, 16 logic processor)	Accuracy Loss Segmentation accuracy Specificity	0.9847 0.0378 0.20 0.9953
Rad 2018 [78]	PA radiographs [28]	*NA** They extracted 155 teeth using segmentation	Inclusion: Age from 25 to 35	*NA*	By a dental examiner	*NA*	*NA*	MLP	*NA*	*NA*	Accuracy on image Accuracy on segments	0.9085 0.98

Abbreviations: **CBCT,** Cone-beam computed tomography; **ICDAS,** International Caries Detection and Assessment System; **GPU,** Graphics Processing Unit; **NA,** Not Available; **mIoU,** mean Intersection over Union; **MCC,** Matthews Correlation Coefficient; **MLP,** Multi-Layer Perceptron; **PA,** Peri-Apical; **R-CNN,** Region Based Convolutional Neural Networks; **ROI,** Region of Interes.

* If only two numbers were written in this column usaully it means Train/Test splites.

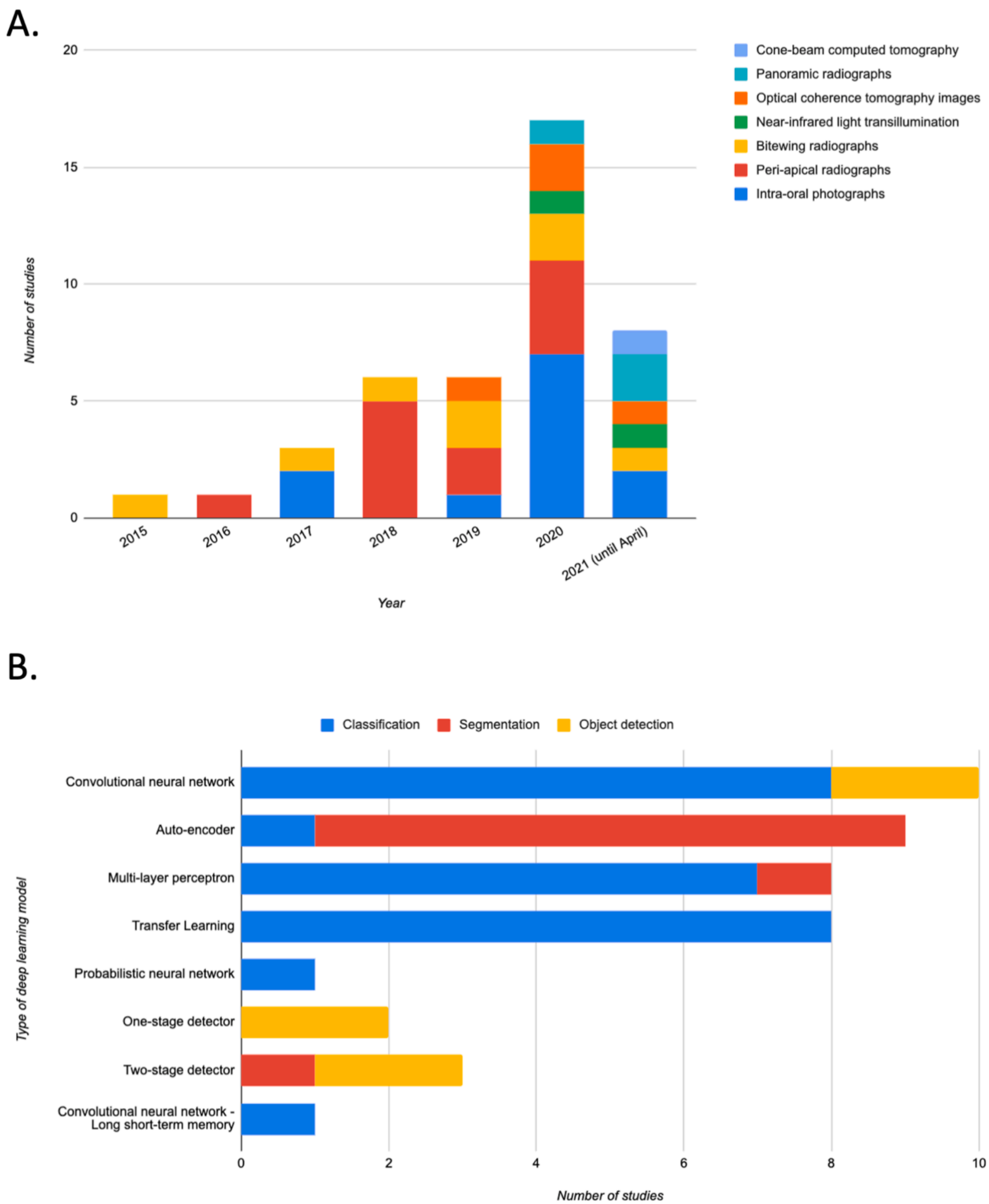

Fig. 2. Included studies; A) Number of deep learning studies for caries detection and image type employed. B) Tasks and modeling approaches involved (*n*=studies).

studies, the level of the performance estimates was heterogeneous across studies). Fourth, one of the most critical shortcomings of the included studies was the lack of detail about the validation and calibration of the annotators and how the lack of inter-annotator agreement was handled. Generally, poor reporting should be addressed by future studies by adhering to minimum standards, e.g., STARD-AI; for diagnostic studies using AI models [36], CLAIM; a checklist for AI in medical imaging, [37], and recent reporting guidelines for dental AI [38]. Fifth, we found a wide range of imagery to be employed for deep learning tasks. Some, like near-infrared light transillumination or optical coherence tomography, are clinically uncommon, and interpretation of such images may not be accurate in the hands of inexperienced examiners [18,39–41]. The latter two types of imagery, however, are clinically promising, as they show high accuracy and do not require ionizing radiation, which is why deep learning seems useful to employ here overcoming the described “experience gap”. Last, it needs highlighting that nearly all included studies only determined the accuracy of deep learning, in a few cases comparing it against the standard of care (dentists without deep learning). Notably, the decisions derived from the diagnostic process are – eventually – more relevant to patients and payers alike, and it remains not understood if deep learning assistance will improve only accuracy, or also decision making and long-term treatment effectiveness and efficiency [42,43]. Ideally, this impact and true usefulness of deep learning for caries detection should be explored in randomized, practice-based settings, and should also reflect aspects like lesion activity as an additional decision point (currently not at all considered by any of the included studies).

This study comes with a number of strengths and limitations. First, it

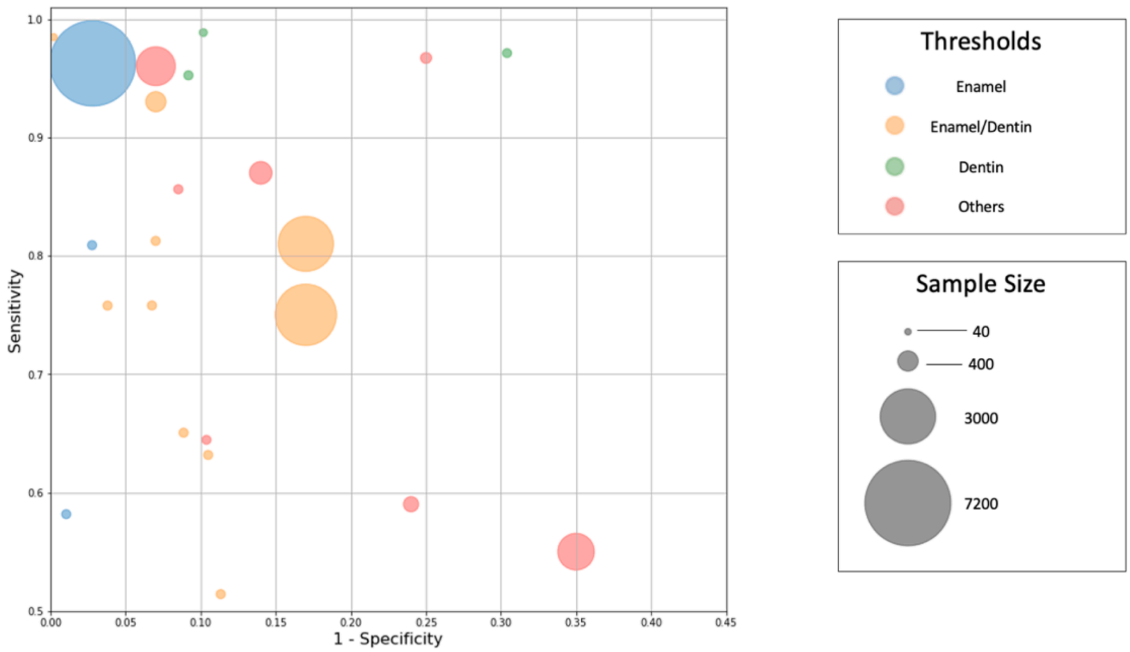

Fig. 3. The various reported sensitivity and specificity of classification studies by sample size. Most of the included studies reported medium to high sensitivity and high specificity. The diameter of the circles represents the size of the dataset.

comprehensively and systematically assessed studies on deep learning for caries detection, and hence allows to (at least narratively) synthesize and contrast them. Second, and as a limitation, our search was limited in its time horizon (which has been justified) and scope (only deep learning has been considered, while alternative image analysis methods have been used previously). Given that we yielded a large and diverse body of evidence, these restrictions seemed justified. Third, we had planned to conduct a meta-analysis, but given the heterogeneity and, more important, limited quality of reporting, we could not extract sufficient detail from most studies to perform this analysis (even if heterogeneity allowed).

5. Conclusions

An increasing number of studies investigated caries detection using deep learning, with a diverse type of architectures being employed. Reported accuracy seems promising, while study and reporting quality are currently low. Future should critically determine the reference test and rely on a comparable, broad and clinically meaningful outcome set.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jdent.2022.104115.

References

[1] T.S. Toh, F. Dondelinger, D. Wang, Looking beyond the hype: applied AI and machine learning in translational medicine, EBioMed. 47 (2019) 607–615, https://doi.org/10.1016/j.ebiom.2019.08.027.

[2] H. Mohammad-Rahimi, M. Nadimi, M.H. Rohban, E. Shamsoddin, V.Y. Lee, S. R. Motamedian, Machine learning and orthodontics, current trends and the future opportunities: a scoping review, Am. J. Orthod. Dentofacial Orthop. 160 (2021) 170–192, https://doi.org/10.1016/j.ajodo.2021.02.013.

[3] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W. M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005.

[4] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117, https://doi.org/10.1016/j.neunet.2014.09.003.

[5] D. Shen, G. Wu, H.-I. Suk, Deep Learning in Medical Image Analysis, Annu. Rev. Biomed. Eng. 19 (2017) 221–248, https://doi.org/10.1146/annurev-bioeng-071516-044442.

[6] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing 234 (2017) 11–26, https://doi.org/10.1016/j.neucom.2016.12.038.

[7] A.P. Malik, M. Pathania, V.K. Rathaur, Overview of artificial intelligence in medicine, J. Family Med. Prim. Care 8 (2019) 2328–2331, https://doi.org/10.4103/jfmpc.jfmpc_440_19.

[8] F. Schwendicke, T. Golla, M. Dreher, J. Krois, Convolutional neural networks for dental image diagnostics: a scoping review, J. Dent. 91 (2019), 103226, https://doi.org/10.1016/j.jdent.2019.103226.

[9] S.B. Khanagar, A. Al-ehaideb, P.C. Maganur, S. Vishwanathaiah, S. Patil, H. A. Baeshen, S.C. Sarode, S. Bhandi, Developments, application, and performance of artificial intelligence in dentistry–A systematic review, J. Dent. Sci. 16 (2021) 508–522, https://doi.org/10.1016/j.jds.2020.06.019.

[10] M.-J. Kim, Y. Liu, S.H. Oh, H.-W. Ahn, S.-H. Kim, G. Nelson, Automatic cephalometric landmark identification system based on the multi-stage convolutional neural networks with CBCT combination images, Sensors 21 (2021) 505, https://doi.org/10.3390/s21020505.

[11] E. Bernabe, W. Marcenes, C.R. Hernandez, J. Bailey, L.G. Abreu, V. Alipour, S. Amini, J. Arabloo, Z. Arefi, A. Arora, M.A. Ayanore, T.W. Bärnighausen, A. Bijani, D.Y. Cho, D.T. Chu, C.S. Crowe, G.T. Demoz, D.G. Demsie, Z.S. Dibaji Forooshani, M. Du, M. El Tantawi, F. Fischer, M.O. Folayan, N.D. Futran, Y.C. D. Geramo, A. Haj-Mirzaian, N. Hariyani, A. Hasanzadeh, S. Hassanipour, S.I. Hay, M.K. Hole, S. Hostiuc, M.D. Ilic, S.L. James, R. Kalhor, L. Kemmer, M. Keramati, Y. S. Khader, S. Kisa, A. Kisa, A. Koyanagi, R. Lalloo, Q. Le Nguyen, S.D. London, N. D. Manohar, B.B. Massenburg, M.R. Mathur, H.G. Meles, T. Mestrovic, A. Mohammadian-Hafshejani, R. Mohammadpourhodki, A.H. Mokdad, S. D. Morrison, J. Nazari, T.H. Nguyen, C.T. Nguyen, M.R. Nixon, T.O. Olagunju, K. Pakshir, M. Pathak, N. Rabiee, A. Rafiei, K. Ramezanzadeh, M.J. Rios-Blancas, E. M. Roro, S. Sabour, A.M. Samy, M. Sawhney, F. Schwendicke, F. Shaahmadi, M. A. Shaikh, C. Stein, M.R. Tovani-Palone, B.X. Tran, B. Unnikrishnan, G.T. Vu, A. Vukovic, T.S.S. Warouw, Z. Zaidi, Z.J. Zhang, N.J. Kassebaum, Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: a systematic analysis for the global burden of disease 2017 study, J. Dent. Res. 99 (2020) 362–373, https://doi.org/10.1177/0022034520908533.

[12] J. Gomez, Detection and diagnosis of the early caries lesion, BMC Oral Health 15 (2015) 1–7, https://doi.org/10.1186/1472-6831-15-S1-S3.

[13] Z. Metzger, D.G. Colson, P. Bown, T. Weihard, I. Baresel, T. Nolting, Reflected near-infrared light versus bite-wing radiography for the detection of proximal caries: a multicenter prospective clinical study conducted in private practices, J. Dent. 116 (2022), 103861, https://doi.org/10.1016/j.jdent.2021.103861.

[14] S. Michou, C. Vannahme, A. Bakhshandeh, K.R. Ekstrand, A.R. Benetti, Intraoral scanner featuring transillumination for proximal caries detection. An in vitro

validation study on permanent posterior teeth, J. Dent. 116 (2022), 103841, https://doi.org/10.1016/j.jdent.2021.103841.
[15] J.-.H. Lee, D.-.H. Kim, S.-.N. Jeong, S.-.H. Choi, Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm, J. Dent. 77 (2018) 106–111, https://doi.org/10.1016/j.jdent.2018.07.015.
[16] C. Deery, Caries detection and diagnosis, sealants and management of the possibly carious fissure, Br. Dent. J. 214 (2013) 551–557, https://doi.org/10.1038/sj.bdj.2013.525.
[17] F. Schwendicke, M. Tzschoppe, S. Paris, Radiographic caries detection: a systematic review and meta-analysis, J. Dent. 43 (2015) 924–933, https://doi.org/10.1016/j.jdent.2015.02.009.
[18] F. Schwendicke, K. Elhennawy, S. Paris, P. Friebertshäuser, J. Krois, Deep learning for caries lesion detection in near-infrared light transillumination images: a pilot study, J. Dent. 92 (2020), 103260, https://doi.org/10.1016/j.jdent.2019.103260.
[19] H. Salehi, M. Barchini, Q. Chen, M. Mahdian, Toward development of automated grading system for carious lesions classification using deep learning and OCT imaging, in: Proc. SPIE 11600, Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging 11600, 2021, pp. 245–252, https://doi.org/10.1117/12.2581318.
[20] H.S. Salehi, M. Barchini, M. Mahdian, Optimization methods for deep neural networks classifying OCT images to detect dental caries, in: Proc. SPIE 11217, Lasers in Dentistry XXVI 11217, 2020, pp. 53–61, https://doi.org/10.1117/12.2545421.
[21] M. Prados-Privado, J. García Villalón, C.H. Martínez-Martínez, C. Ivorra, J. C. Prados-Frutos, Dental caries diagnosis and detection using neural networks: a systematic review, J. Clin. Med. 9 (2020) 3579, https://doi.org/10.3390/jcm9113579.
[22] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini, F. Schürmann, I. Krejci, H. Markram, Caries detection with near-infrared transillumination using deep learning, J. Dent. Res. 98 (2019) 1227–1233, https://doi.org/10.1177/0022034519871884.
[23] M.M. Srivastava, P. Kumar, L. Pradhan, S. Varadarajan, Detection of tooth caries in bitewing radiographs using deep learning, arXiv preprint arXiv:1711.07312 (2017).
[24] M.D.F. McInnes, D. Moher, B.D. Thombs, T.A. McGrath, P.M. Bossuyt, T. Clifford, J. F. Cohen, J.J. Deeks, C. Gatsonis, L. Hooft, H.A. Hunt, C.J. Hyde, D.A. Korevaar, M. M.G. Leeflang, P. Macaskill, J.B. Reitsma, R. Rodin, A.W.S. Rutjes, J.P. Salameh, A. Stevens, Y. Takwoingi, M. Tonelli, L. Weeks, P. Whiting, B.H. Willis, Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement, J. Am. Med. Assoc. 319 (2018) 388–396, https://doi.org/10.1001/jama.2017.19163.
[25] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (NIPS) 25 (2012) 1097–1105.
[26] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M. M. Leeflang, J.A. Sterne, P.M. Bossuyt, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (2011) 529–536, https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
[27] C.W. Wang, C.T. Huang, J.H. Lee, C.H. Li, S.W. Chang, M.J. Siao, T.M. Lai, B. Ibragimov, T. Vrtovec, O. Ronneberger, P. Fischer, T.F. Cootes, C. Lindner, A benchmark for comparison of dental radiography analysis algorithms, Med. Image Anal. 31 (2016) 63–76, https://doi.org/10.1016/j.media.2016.02.004.
[28] A.E. Rad, M.S.M. Rahim, A. Rehman, T. Saba, Digital dental X-ray database for caries screening, 3D Res 7 (2016) 1–5, https://doi.org/10.1007/s13319-016-0096-5.
[29] Kaggle, Teeth_dataset, Kaggle, https://www.kaggle.com/pushkar34/teeth-dataset/version/1, 2020 (Accessed 09-Oct-2020).
[30] C. Wang, H. Qin, G. Lai, G. Zheng, H. Xiang, J. Wang, D. Zhang, Automated classification of dual channel dental imaging of auto-fluorescence and white lightby convolutional neural networks, J. Innovat. Opt. Health Sci. 13 (2020), 2050014, https://doi.org/10.1142/S1793545820500145.
[31] Y.-.P. Huang, S.-.Y. Lee, An effective and reliable methodology for deep machine learning application in caries detection, *medRxiv:2021.05.04.21256502* (2021).
[32] M.J. Willemink, W.A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R.M. Summers, D.L. Rubin, M.P. Lungren, Preparing medical imaging data for machine learning, Radiology 295 (2020) 4–15, https://doi.org/10.1148/radiol.2020192224.
[33] J. Lee, K.W. Kim, S.H. Choi, J. Huh, S.H. Park, Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part ii. statistical methods of meta-analysis, Korean J. Radiol. 16 (2015) 1188–1196, https://doi.org/10.3348/kjr.2015.16.6.1188.
[34] J.B. Reitsma, A.S. Glas, A.W. Rutjes, R.J. Scholten, P.M. Bossuyt, A.H. Zwinderman, Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, J. Clin. Epidemiol. 58 (2005) 982–990, https://doi.org/10.1016/j.jclinepi.2005.02.022.
[35] J. Dinnes, J. Deeks, J. Kirby, P. Roderick, A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy, Health Technol. Assess. 9 (2005) 1–113, https://doi.org/10.3310/hta9120.
[36] V. Sounderajah, H. Ashrafian, R. Aggarwal, J. De Fauw, A.K. Denniston, F. Greaves, A. Karthikesalingam, D. King, X. Liu, S.R. Markar, M.D.F. McInnes, T. Panch, J. Pearson-Stuttard, D.S.W. Ting, R.M. Golub, D. Moher, P.M. Bossuyt, A. Darzi, Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI Steering Group, Nat. Med. 26 (2020) 807–808, https://doi.org/10.1038/s41591-020-0941-1.
[37] J. Mongan, L. Moy, C.E. Kahn, Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a Guide for authors and reviewers, Radiol.: Artif. Intell. 2 (2020), e200029, https://doi.org/10.1148/ryai.2020200029.
[38] F. Schwendicke, T. Singh, J.H. Lee, R. Gaudin, A. Chaurasia, T. Wiegand, S. Uribe, J. Krois, Artificial intelligence in dental research: checklist for authors, reviewers, readers, J. Dent. 107 (2021), 103610, https://doi.org/10.1016/j.jdent.2021.103610.
[39] M.I.G. Ortiz, C. de Melo Alencar, B.L.F. De Paula, M.B. Magno, L.C. Maia, C. M. Silva, Accuracy of near-infrared light transillumination (NILT) compared to bitewing radiograph for detection of interproximal caries in the permanent dentition: a systematic review and meta-analysis, J. Dent. 98 (2020), 103351, https://doi.org/10.1016/j.jdent.2020.103351.
[40] A. Holtkamp, K. Elhennawy, J.E. Cejudo Grano de Oro, J. Krois, S. Paris, F. Schwendicke, Generalizability of deep learning models for caries detection in near-infrared light transillumination images, J. Clin. Med. 10 (2021) 961, https://doi.org/10.3390/jcm10050961.
[41] H. Schneider, M. Ahrens, M. Strumpski, C. Rüger, M. Häfer, G. Hüttmann, D. Theisen-Kunde, H. Schulz-Hildebrandt, R. Haak, An intraoral OCT Probe to enhanced detection of approximal carious lesions and assessment of restorations, J. Clin. Med. 9 (2020) 3257, https://doi.org/10.3390/jcm9103257.
[42] F. Schwendicke, J.G. Rossi, G. Göstemeyer, K. Elhennawy, A.G. Cantu, R. Gaudin, A. Chaurasia, S. Gehrung, J. Krois, Cost-effectiveness of artificial intelligence for proximal caries detection, J. Dent. Res. 100 (2021) 369–376, https://doi.org/10.1177/0022034520972335.
[43] V. Baelum, H. Hintze, A. Wenzel, B. Danielsen, B. Nyvad, Implications of caries diagnostic strategies for clinical management decisions, Community Dent. Oral Epidemiol. 40 (2012) 257–266, https://doi.org/10.1111/j.1600-0528.2011.00655.x.
[44] L.M. Leo, T.K. Reddy, Learning compact and discriminative hybrid neural network for dental caries classification, Microprocess Microsyst. 82 (2021), 103836, https://doi.org/10.1016/j.micpro.2021.103836.
[45] L. Megalan Leo, T. Kalpalatha Reddy, Dental caries classification system using deep learning based convolutional neural network, J. Comput. Theor. Nanosc. 17 (2020) 4660–4665, https://doi.org/10.1166/jctn.2020.9295.
[46] P. Tripathi, C. Malathy, M. Prabhakaran, Genetic algorithms based approach for dental caries detection using back propagation neural network, Int. J. Recent Tech. Eng. 8 (2019) 317–319.
[47] A. Sonavane, R. Yadav, A. Khamparia, Dental cavity classification of using convolutional neural network, IOP Conf. Ser.: Mater. Sci. Eng. 1022 (2021) 12116, https://doi.org/10.1088/1757-899X/1022/1/012116.
[48] P. Singh, P. Sehgal, G.V Black dental caries classification and preparation technique using optimal CNN-LSTM classifier, Multimed. Tools Appl. 80 (2021) 5255–5272, https://doi.org/10.1007/s11042-020-09891-6.
[49] P. Singh, P. Sehgal, Decision support system for black classification of dental images using GIST descriptors, Adv. Intell. Syst. Comput. 1082 (2020) 343–352, https://doi.org/10.1007/978-981-15-1081-6_29.
[50] A.A. Guijarro-Rodríguez, P.M. Witt-Rodríguez, L.J. Cevallos-Torres, S.F. Contreras-Puco, M.C. Ortiz-Zambrano, D.E. Torres-Martínez, Image segmentation techniques application for the diagnosis of dental caries, Adv. Intell. Syst. Comput. 1066 (2020) 312–322, https://doi.org/10.1007/978-3-030-32022-5_30.
[51] S.A. Prajapati, R. Nagaraj, S. Mitra, Classification of dental diseases using CNN and transfer learning, in: 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), 2017, pp. 70–74, https://doi.org/10.1109/ISCBI.2017.8053547.
[52] H.S. Salehi, M. Mahdian, M.M. Murshid, S. Judex, A. Tadinada, Deep learning-based quantitative analysis of dental caries using optical coherence tomography: an ex vivo study, in: Proc. SPIE 10857, Lasers in Dentistry XXVI 10857, 2019, pp. 39–46, https://doi.org/10.1117/12.2510076.
[53] S. Riyadi, S. Mayanti, C. Damarjati, S. Puspita, Deep learning for pixel-based edge models classification of tertiary dentine images, in: Proceedings of the 2020 The 3rd International Conference on Information Science and System, 2020, pp. 57–60, https://doi.org/10.1145/3388176.3388197.
[54] V. Geetha, K.S. Aprameya, D.M. Hinduja, Dental caries diagnosis in digital radiographs using back-propagation neural network, Health Inf. Sci. Syst. 8 (2020) 8, https://doi.org/10.1007/s13755-019-0096-y.
[55] S. Patil, V. Kulkarni, A. Bhise, Intelligent system with dragonfly optimisation for caries detection, IET Image Process 13 (2019) 429–439, https://doi.org/10.1049/iet-ipr.2018.5442.
[56] M. Sornam, M. Prabhakaran, Logit-based artificial bee colony optimization (LB-ABC) approach for dental caries classification using a back propagation neural network, in: A. Krishna, K. Srikantaiah, C. Naveena (Eds.), Integrated Intelligent Computing, Communication and Security, Studies in Computational Intelligence, 2019, pp. 79–91, https://doi.org/10.1007/978-981-10-8797-4_9.
[57] S. Patil, V. Kulkarni, A. Bhise, Caries detection using multidimensional projection and neural network, Int. *J. Knowl. Base Intell. Eng. Syst.* 22 (2018) 155–166, https://doi.org/10.3233/KES-180381.
[58] S. Patil, V. Kulkarni, A. Bhise, Caries detection with the aid of multilinear principal component analysis and neural network, in: Proceedings of the 2nd International Conference on Green Computing and Internet of Things, ICGCIoT 2018, 2018, pp. 272–277, https://doi.org/10.1109/ICGCIoT.2018.8753002.
[59] J.H. Lee, D.H. Kim, S.N. Jeong, S.H. Choi, Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm, J. Dent. 77 (2018) 106–111, https://doi.org/10.1016/j.jdent.2018.07.015.
[60] M. Sornam, M. Prabhakaran, A new linear adaptive swarm intelligence approach using back propagation neural network for dental caries classification, in: IEEE International Conference on Power, Control, Signals and Instrumentation

Engineering, ICPCSI 2017, 2018, pp. 2698–2703, https://doi.org/10.1109/ICPCSI.2017.8392208.
[61] M.M. Lakshmi, P. Chitra, Classification of dental cavities from X-ray images using deep CNN algorithm, in: 4th International Conference on Trends in Electronics and Informatics (ICOEI), 2020, pp. 774–779, https://doi.org/10.1109/ICOEI48184.2020.9143013.
[62] S. Vinayahalingam, S. Kempers, L. Limon, D. Deibel, T. Maal, S. Bergé, T. Xi, M. Hanisch, The automatic detection of caries in third molars on panoramic radiographs using deep learning: a pilot study, Res. Sq. (2021), https://doi.org/10.21203/rs.3.rs-379636/v1.
[63] A. Haghanifar, M.M. Majdabadi, S.-.B. Ko, PaXNet: dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier, arXiv preprint arXiv:2012.13666 (2020).
[64] X. Zhang, Y. Liang, W. Li, C. Liu, D. Gu, W. Sun, L. Miao, Development and evaluation of deep learning for screening dental caries from oral photographs, Oral Dis. 28 (2021) 173–181, https://doi.org/10.1111/odi.13735.
[65] A. Javid, U. Rashid, A.S. Khattak, Marking early lesions in labial colored dental images using a transfer learning approach, in: IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1–5, https://doi.org/10.1109/INMIC50486.2020.9318173.
[66] H. Yu, Z. Lin, Y. Liu, J. Su, B. Chen, G. Lu, A new technique for diagnosis of dental caries on the children's first permanent molar, IEEE Access 8 (2020) 185776–185785, https://doi.org/10.1109/ACCESS.2020.3029454.
[67] Y. Zhang, H. Liao, J. Xiao, N.A. Jallad, O. Ly-Mapes, J. Luo, A smartphone-based system for real-time early childhood caries diagnosis, in: In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. ASMUS PIPPI 2020, Cham, Springer, 2020, pp. 233–242, https://doi.org/10.1007/978-3-030-60334-2_23. Lecture Notes in Computer Science, vol 12437.
[68] J. Choi, H. Eun, C. Kim, Boosting proximal dental caries detection via combination of variational methods and convolutional neural network, J. Sign. Process. Syst. 90 (2018) 87–97, https://doi.org/10.1007/s11265-016-1214-6.
[69] A.G. Cantu, S. Gehrung, J. Krois, A. Chaurasia, J.G. Rossi, R. Gaudin, K. Elhennawy, F. Schwendicke, Detecting caries lesions of different radiographic extension on bitewings using deep learning, J. Dent. 100 (2020), 103425, https://doi.org/10.1016/j.jdent.2020.103425.
[70] P. Kumar, M.M. Srivastava, Example mining for incremental learning in medical imaging, in: Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, 2018, pp. 48–51, https://doi.org/10.1109/SSCI.2018.8628895.
[71] J. Yun, T. Ning, Z. Hai, P. Tingting, Bitewing radiography semantic segmentation base on conditional generative adversarial nets, arXiv preprint arXiv:1802.02571 (2018).
[72] O. Ronneberger, P. Fischer, T. Brox, Dental X-ray image segmentation using a U-shaped deep convolutional network, in: International Symposium on Biomedical Imaging, 2015, pp. 1–13, 1.
[73] M. Ezhov, M. Gusarev, M. Golitsyna, J. Yates, E. Kushnerev, D. Tamimi, S. Aksoy, E. Shumilov, A. Sanders, K. Orhan, Development and validation of a Cbct-based artificial intelligence system for accurate diagnoses of dental diseases, Res. Sq. (2021), https://doi.org/10.21203/rs.3.rs-303329/v1.
[74] K. Moutselos, E. Berdouses, C. Oulis, I. Maglogiannis, Recognizing occlusal caries in dental intraoral images using deep learning, in: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 1617–1620, https://doi.org/10.1109/EMBC.2019.8856553.
[75] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini, F. Schürmann, I. Krejci, H. Markram, Caries detection with near-infrared transillumination using deep learning, J. Dent. Res. 98 (2019) 1227–1233, https://doi.org/10.1177/0022034519871884.
[76] H.A. Khan, M.A. Haider, H.A. Ansari, H. Ishaq, A. Kiyani, K. Sohail, M. Muhammad, S.A. Khurram, Automated feature detection in dental periapical radiographs by using deep learning, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 131 (2021) 711–720, https://doi.org/10.1016/j.oooo.2020.08.024.
[77] Y.J. Jung, M.J. Kim, Deeplab v3+ based automatic diagnosis model for dental X-ray: preliminary study, J. Magn. 25 (2020) 632–638, https://doi.org/10.4283/JMAG.2020.25.4.632.
[78] A.E. Rad, M.S.M. Rahim, H. Kolivand, A. Norouzi, Automatic computer-aided caries detection from dental x-ray images using intelligent level set, Multimed. Tools Appl. 77 (2018) 28843–28862, https://doi.org/10.1007/s11042-018-6035-0.